Marseille, 11–16 May 2020

Developing Resources for Automated Speech Processing of Quebec French

Mélanie Lancien[1,2], **Marie-Hélène Côté**[1], **Brigitte Bigi**[3]
Université de Lausanne[1], Université du Québec à Chicoutimi[2], LPL-CNRS-Aix Marseille Univ[3]
Lausanne (Suisse), Chicoutimi (Canada), Aix-en-Provence (France)
melanie.lancien@unil.ch, marie-helene.cote@unil.ch, brigitte.bigi@lpl-aix.fr

Abstract

The analysis of the structure of speech nearly always rests on the alignment of the speech recording with a phonetic transcription. Nowadays several tools can perform this speech segmentation automatically. However, none of them carries out the automatic segmentation of Quebec French (QF hereafter) in a proper way. Contrary to what could be assumed, the acoustics and phonotactics of QF differs widely from that of France French (FF hereafter). To adequately segment QF, features like diphthongization of long vowels and affrication of coronal stops have to be taken into account. Thus acoustic models for automatic segmentation must be trained on speech samples exhibiting those phenomena. Dictionaries and lexicons must also be adapted and integrate differences in lexical units (such as very frequent words in QF that are not used in FF) and in the phonology of QF (such as the existence of tense and lax high vowels in QF but not in FF). This paper presents the development of linguistic resources to be included into the SPPAS software tool in order to get Text normalization, Phonetization, Alignment and Syllabification. We adapted the existing French lexicon and developed a QF-specific pronunciation dictionary. We then created an acoustic model from the existing ones and adapted it with 5 minutes of manually time-aligned data. These new resources are all freely distributed with SPPAS version 2.7; they perform the full process of speech segmentation in Quebec French.

Keywords: Speech segmentation, Quebec French, resources, dictionary, phonetics

1. Introduction

The physical characteristics of speech can be observed in the dynamics of the production of sounds. One of the main goals of phonetic studies is to capture, describe and explain the physical characteristics of linguistic units. First, such studies mostly consist in determining the relevant units and aligning them with segments of the audio signal. Then they measure parameters of the units (e.g. their duration) and, finally, they perform some descriptive statistical analyses. Annotating the units is thus crucially important. In this context, the required annotations consist in performing a phonetic alignment, i.e. labelling and segmenting phonemes. Labelling determines the sounds of a transcription of a speech recording. Segmenting consists in pairing these sounds with time-stamps to indicate the localization and thus the length of the sounds in the speech recording. Linguists need tools for the analysis of these features of speech sounds. The most popular software tool to perform these annotations manually is Praat (Boersma and Weenink, 2018), but other tools combine ease of use with a broad range of features like Phonedit (Teston et al., 1999) or AnnotationPro (Klessa et al., 2013).

Several tools for automatic phonetization and segmentation are already available for the French language. However, those are usually trained and designed for France French (Schiel, 1999; Goldman, 2011; Christodoulides, 2018) or based on a dictionary following the phonetic patterns of standard French for the phonetization of the orthographic forms (McAuliffe et al., 2017).

When it comes to processing regional varieties of French such as Quebec French, which works far differently from France French from a phonological and phonetic point of view, no suitable tool for automatic processing is available for the research community. Quebec French (QF hereafter) undergoes a lot of processes, such as stop affrication, high vowel laxing, or long vowel diphthongization, that affect the phonetic realization of its phonemes. None of those (except maybe a stylistic use of stop affrication (Trimaille et al., 2012)) are relevant in standard (France) French (FF hereafter), therefore when confronted with a diphthongized vowel, automatic speech processing tools can't establish the right segmentation for the phone (since models were not trained on speech exhibiting this phenomenon).

This paper describes the linguistic resources we created to perform the automatic speech segmentation of QF using SPPAS - the automatic annotation and analysis of speech (Bigi, 2015). This system divides this task into three subtasks, namely text normalization, phonetization, alignment and syllabification. Each of these automatic annotations requires a linguistic resource. A list of words is required for the text normalization task; as SPPAS already had one, we didn't have to construct such a vocabulary but to adapt it. We had to establish a phonological inventory and to create a pronunciation dictionary to be included into SPPAS to perform phonetization. We then built an acoustic model from both the already existing models included into SPPAS and a few minutes of QF speech we manually time-aligned. Finally, we adapted the existing syllabification configuration file for FF to be suitable for QF. These newly created resources are freely included into version 2.7 of SPPAS, allowing it to perform the automatic speech segmentation of QF.

2. Automatic speech segmentation: overview of the method

In recent years, the SPPAS software tool (Bigi, 2015) has been developed to automatically produce annotations and to analyze annotated data. SPPAS is multi-platform (Linux, MacOS and Windows) and open source. As a main functionality, it performs speech segmentation of recorded audio (speech) files and its orthographic transcription. Speech segmentation is the process of identifying the phonemes and their boundaries in speech recordings. Creating an automatic speech segmentation system commonly includes the development of both a text normalization system and a grapheme-to-phoneme system, the data preparation (large amount of recordings and their manual labelling), the building of an acoustic model, and finally the use of an automatic segmenter. The SPPAS system reduces significantly the list of such requirements and allows speech segmentation for a new language without modifying the internal algorithms: only the linguistic resources need to be built and added into the appropriate directories of its package.

2.1. Text Normalization

Text normalization is an automatic process that consists in formatting the orthographic transcription. SPPAS can perform such normalization from a standard transcription but it also supports an enriched orthographic transcription. In spontaneous speech numerous phenomena occur, such as hesitations, repetitions, non-standard elisions, reduction phenomena, truncated words, and, more generally, non-standard pronunciations. Events like laughter, noises and filled pauses are also very frequently observed in spontaneous speech (Bigi and Meunier, 2018). The Enriched Orthographic Transcription convention (EOT) of the SPPAS software tool includes the possibility to add the following phenomena into the manual transcription:

- a breath, a cough or an unintelligible segment is noted '*';
- laughter is noted '@';
- a short pause is noted '+';
- a broken word is noted with a '-'at the end of the token string;
- an elision is mentioned in parentheses, like thi(s);
- a specific pronunciation is noted with brackets like this [example, eczap];
- an unexpected liaison is surrounded by '=';

The Text Normalization of SPPAS can derive two normalized texts from the EOT (Bigi et al., 2012). The standard one contains the standard form of words and the faked one contains a phonetic-form of words. For example, the phrase " thi(s) [example, eczap] " has the standard form "this example" and its faked-form is "thi eczap". The first one is human-readable and relevant for syntactic or lexical analyses but the second one provides a better grapheme-to-phoneme conversion and so a better time-alignment with the audio stream.
To perform such text normalization for a given language, a list of words is required: SPPAS implements a language-independent algorithm (Bigi, 2014). The number-to-letter system is actually the only language-dependent module in the system and it doesn't need to be modified: QF is similar to FF - contrary to Belgium French, for example, which is slightly different (70 is read *septante* /sɛptɑ̃t/ instead of *soixante-dix* /swasɑ̃tdis/, for instance). In fact, the number-to-letter system of QF can be different for numbers from 62 to 79. They can either be pronounced in the standard FF way (62 is *soixante-deux* /swasɑ̃tdø/) or differ from it (62 is *soixante-et-deux* /swasɑ̃tedø/). This is a speaker-dependent variant that we can't take into account in the automatic system. In case a number is not pronounced in the standard way, the number can be written like any other specific pronunciation[1], so "62" is transcribed "[62, soixante_et_deux]".
So, integrating QF into SPPAS only required the creation of a lexicon or the adaptation of the existing one for FF.

2.2. Phonetization

Phonetization is the process of representing sounds by phonetic symbols. There are two general ways to construct a phonetization process: rule-based systems and dictionary-based solutions, which consist in storing a maximum of phonological knowledge in a lexicon. Contrary to most other tools, SPPAS implements the latter: it doesn't require any training stage so it's well suited for under-resourced languages (Bigi, 2016).
The SPPAS program for the phonetization of the normalized orthographic transcription produces a phonetic transcription based on a phonetic dictionary. When a word is missing from the dictionary, an empirical method based on the use of the existing dictionary is implemented to produce a set of possible pronunciations. When a word can be pronounced in several ways, all pronunciation variants are expected to be included into the pronunciation dictionary. At this stage of the process, SPPAS generates all possibilities and the appropriate variant is determined during alignment.
An important step in the construction of the speech segmentation system is then to build the pronunciation dictionary, where each word in the vocabulary is expanded into its constituent phones.

2.3. Alignment

To perform the alignment task, SPPAS operates as a wrapper either for HVite command of HTK toolkit or for julius command of the Julius CSR engine, which is the default. In any case, a finite state grammar describing sentence patterns to be recognized and an acoustic model are needed.
The grammar is automatically created from the result of the phonetization when a new utterance is decoded. The aligner - HVite or Julius, performs a two-stage process in which the first pass determines the appropriate path in the grammar and so the labelling. The second pass consists in finding the time-stamps of each label.
The required acoustic model has to be created and included in the resources of the software. In SPPAS, phoneme-based Hidden Markov Models (commonly named "monophone" HMM) were chosen and saved in HTK-ASCII format. The HMM approach is used to model phonemes in individual HMM with a 5-state model with a left-to-right topology with self-loops and no transitions skipping over states. The feature vector that represents the distinctive properties of the phoneme is designed to be of length 39, consisting of 12 mel-cepstrum coefficients and energy component, and additionally their delta and acceleration coefficients. This model is able to handle new data robustly in different speech

[1] See the transcription convention for details: https://www.ortolang.fr/market/item/sldr000873

styles and can predict efficiently which phone was uttered (Bigi and Meunier, 2018). This kind of HMM model is also preferred because it is language independent. Actually, an acoustic model comprises each HMM representing all the phonemes of a given language. Some events that frequently occur in speech like noises, laughter or filled pauses can also be represented individually by an HMM and included in the model, like in the French acoustic model available in the resources of SPPAS.

To construct an acoustic model for a given language, it is then possible to train a new model from a large amount of data (if available); or to group together the required HMMs picked up from other models (when available), and such a model can eventually be adapted to the given language with a small set of data (about 3-5 minutes). This second approach was previously successfully applied on the low-resourced language Naija (Nigerian Pidgin) (Bigi et al., 2017). We thus applied the same methodology for Quebec French.

2.4. Syllabification

Like in the previous release of the software, syllabification is performed from time-aligned phonemes (Bigi et al., 2010). This phoneme-to-syllable segmentation is a rule-based system based on 2 main principles : first, a syllable contains a vowel, and only one; second, a pause is a syllable boundary. Rules are established to distribute consonants between 2 vowels.

syllabification is implemented thanks to a rule-based system for which the list of phonemes and the rules are fixed in an external configuration file. To get an automatic syllabification of QF, it was only required to add the phonemes and their class into the configuration file of FF.

3. A lexicon of QF

A vocabulary for FF is already available in the resources of SPPAS. The word items we added for the QF lexicon were either (for the most obvious ones) picked by a native speaker (our 2nd author) or extracted from the manual orthographical transcription of 143 hours of spontaneous speech from the PFC-Quebec corpus. This corpus was recorded as part of the PFC project (Durand et al., 2002). We focus here on the PFC sub-corpus recorded in Quebec and neighboring regions, composed of 410 hours of speech (90 hours of reading, 320 hours of conversational speech) from 462 native French speakers from 34 localities (survey points) (Côté, 2014). Recordings were done using different brands of portable wireless recorders with built-in stereo microphones (belonging to the Tascam and Zoom series).

In order to perform Text normalization of QF with SPPAS, users have to select French in the list of available languages.

4. Phonetic description of Quebec French & Pronunciation dictionary

Constructing an automatic system for speech segmentation first requires the definition of the set of phonemes to be used in the pronunciation dictionary. These phonemes will have to be represented in the acoustic model. To do so for QF, we relied mainly on (Côté, 2012) and (Santerre, 1976). This variety of French has 18 consonants, three glides, and between 15 and 23 vowels (depending on the authors). Here we selected:

- the 18 consonants /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ, l, m, n, ʁ, ŋ, ɲ/
- the three glides /ɥ, w, j/
- 20 vowels /i, y, u, ɪ, ʏ, ʊ, e, ɛ, ɜ, ø, œ, ə, a, ɑ, o, ɔ, ɛ̃, œ̃, ɔ̃, ɑ̃/

Regarding the consonants and glides, QF and FF share the same phonemic inventories. However, the QF alveolar stops are subject to affrication rules, /t, d/ being realized /ts, dz/ before high front vowels and glides, and the phoneme /R/ has several allophonic realizations, including apical trills and flaps, which are absent in standard French, and vocalized variants in word-final and preconsonantal position, which are particularly frequent in QF. When it comes to vowels, the two varieties differ markedly. In addition to the four categories /ɪ, ʏ, ʊ, ɜ/ that do not exist in FF, QF also displays diphthongized realizations of long vowels in some word positions, which affect both the quality and the quantity of the vowel. These processes can affect the recognition of the phone in the signal and the way it is segmented.

Our pronunciation dictionary of QF includes more vowels, in comparison with FF. This implies that, for every /i, y, u, ɛ/ in FF, it must be determined whether it corresponds to the same vowel in QF or to the additional /ɪ, ʏ, ʊ, ɜ/. The distribution of each pair - /i/ vs. /ɪ/, /y/ vs. /ʏ/, /u/ vs. /ʊ/ and /ɛ/ vs. /ɜ/ - is partly lexicalized and partly predictable. For the high vowels, the lax variants appear, with some lexical exceptions, in final syllables closed by a consonant other than /R, v, z, ʒ/; the tense high vowels surface everywhere else. The distribution of /ɛ/ vs. /ɜ/ is less predictable. In word-final position, only /ɛ/ surfaces, while /ɜ/ is found systematically before word-final /R/. In final syllables closed by a consonant other than /R/, two word classes are distinguished: the /ɛ/ one (e.g. *faites* 'you-PL make' /fɛt/) and the /ɜ/ one (e.g. *fête* 'party' /fɜt/). The /ɜ/ vowel is usually retained in derived words, where it appears in non-final syllables (e.g. *fêter* 'to party' /fɜte/), but this is not systematic and must be lexically determined (e.g. *maître* 'master' /mɜtʁ/, *maîtrise* 'master's degree' /mɜtʁiz/, but *maîtresse* 'mistress' /mɛtʁɛs/). The distribution of other vowel pairs may also be partly predictable, for instance for the two low vowels /a/ and /ɑ/ in word-final position and before final /ʁ/, where /a/ appears in restricted contexts or specific lexical items. Such specific features of the sound system of QF were taken into account in our pronunciation dictionary.

The dictionary is designed to also include pronunciation variants for words that have a variable surface form. Such cases are either lexically specific (e.g. *fait* 'fact, done' /fɛ/-/fɛt/) or governed by more general rules. These include the deletion of /ə/ in word-initial and final syllables (e.g. *tenu* 'held' /təny/-/tny/; *quatre* /katʁə/-/katʁ/), and the pronunciation of liaison consonants (e.g. *les amis* 'friends' /lezami/). Both processes also concern FF, but they function differently in QF (Côté, 2012). In addition, the deletion of word-final consonants in certain clusters is particularly frequent in QF, including before vowel-initial words, leading to systematic pronunciation variants like *table* ‘table’ /tabl/-/tab/ and *acte* ‘act’ /akt/-/ak/. QF also displays several cases of determiner+preposition or pronoun+auxiliary mergers, which involve the deletion of initial /l/ in the pronouns and determiners *la* and *les* and vowel fusion, as in *dans les* ‘in the.PL’ pronounced /dɑ̃le/-/dɑ̃e/-/dẽ/.

SPPAS	IPA	Description
a	a	open front unrounded
A	ɑ	open back unrounded
2	ø	close-mid front rounded
3	ɜ	open-mid central unrounded
9	œ	open-mid front rounded
i	i	close front unrounded
I	ɪ	near-close front unrounded
e	e	close-mid front unrounded
E	ɛ	open-mid front unrounded
o	o	close-mid back rounded
O	ɔ	open-mid back rounded
u	u	close back rounded
U	ʊ	near-close back rounded
y	y	close front rounded
Y	ʏ	near-close front rounded
@	ə	schwa
A~	ɑ̃	nasal
E~	ɛ̃	nasal
O~	ɔ̃	nasal
U~/	œ̃	nasal

Table 1: List of vowels

The full dictionary, which currently contains 164k forms, is regularly corrected and is planned to be freely distributed, but is currently available on demand. At the time of writing this paper, a version with 40k words is already included into SPPAS. All the sounds are represented with the ASCII-based SAMPA code (Table 1).
In order to perform Phonetization of QF with SPPAS, users have to select Quebec French in the list of available languages.

5. Acoustic model

During the training procedure of an acoustic model, the first step is commonly to initialize the model; this step is often called "bootstrap". It aims to create an HMM prototype for each phoneme using time-aligned data. This step was replaced by the use of phoneme prototypes already available in other languages. In the SPPAS package, ten acoustic models of the same type - i.e. same HMMs definition and same parameters - were trained and are freely distributed with a public license. These phoneme prototypes can be extracted and reused in English, French, Italian, Spanish, Catalan, Polish, Mandarin Chinese, Southern Min, Naija languages.
For QF, we extracted the HMMs mainly from French, but also German, English and Polish. As we can see in Table 2, all consonants except one were extracted from the French model, unlike vowels which are picked up from the four languages. This is mainly a consequence of the fact that the French model of SPPAS A/ represents a or A, O/ represents o or O and U~/ represents e~ or 9~, in SAMPA code[2]. Finally, not all HMMs of the QF phonemes were available in the models and we had to map some of them:

- ã French model for ɑ̃
- œ French model for ɜ
- j French model for ɲ

fra	b d f g H j k l m n p R s S t v w z Z
fra	9 @ 2 e E i o O~ u U~/ y
deu	a O Y N
eng	A I U
pol	E~

Table 2: Inventory of phonemes and language from which the HMMs were extracted

This approach enables the acoustic model to be trained by a small amount of target language speech data (Le et al., 2008), and it has proven its efficiency even with only three minutes of manually time-aligned data in (Bigi et al., 2017). From this set of time-aligned data (at the level of phonemes and/or IPUs), the training procedure implemented into SPPAS is based on the HTK-Toolkit as described in the HTK tutorial. The initial QF acoustic model created from the HMMs of other languages is then adapted to QF with a corpus we created for this purpose.
The initial corpus is the PFC sub-corpus recorded in Quebec and neighboring regions, as described above in the Lexicon section. By now 350 hours of speech have been orthographically transcribed and are planned to be automatically segmented with the SPPAS resources described here. From them about seven minutes of read speech by eight speakers were manually time aligned with Praat (Boersma and Weenink, 2018), yielding 2,716 tokens of the 41 phonemes of QF.

This corpus was divided into both a training subset (273 seconds of speech, 2390 phonemes), from which the acoustic model was adapted, and a test subset (98 seconds of speech, 765 phonemes), in order to evaluate the resulting automatic alignment (Figure 1).
Finally, the HMM representing the noise, the one representing the laughter and the one representing the hesitation ("euh") were all included into the acoustic model in order to time-align spontaneous speech like conversations, interviews, etc. (Bigi and Meunier, 2018).
In order to perform Alignment of QF with SPPAS, users have to select Quebec French in the list of available languages.

6. Syllabification configuration

Only the additional QF vowels were added into the FF configuration file for the automatic syllabification of QF: I, Y,

[2] Extended version of French SAMPA proposed by J.C. Wells at http://www.phon.ucl.ac.uk/home/sampa/french.htm

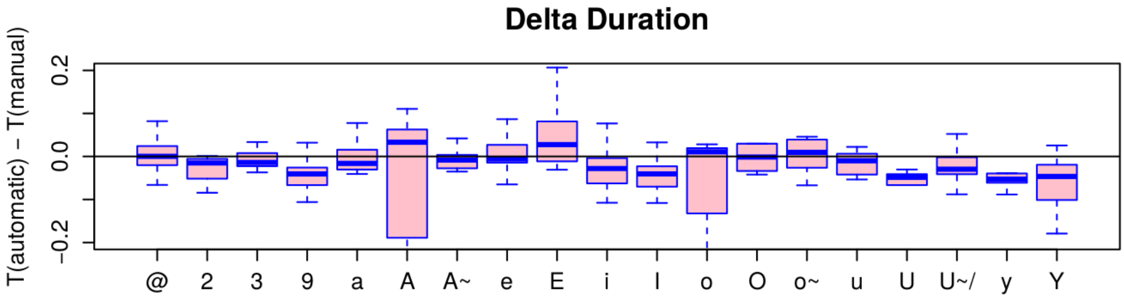

Figure 1: Duration of vowels: a comparison of vowels' durations as obtained with the automatic vs manual alignment (phonemes are represented in SAMPA).

U, 3, A˜. In order to perform Syllabification of QF with SPPAS, users have to select French in the list of available languages.

7. Final outcome

We developed or adapted all the resources required to operate the automatic speech segmentation of QF in SPPAS. Figure 3 is a screenshot of an extract of speech from one of the reading tasks of the speaker COCJG1 (female), from the Hawkesbury survey point of the PFC corpus. The first tier is the orthographic transcription and the last tier is the manual time-alignment of phonemes. All the other tiers were generated automatically. Figure 2 is a screenshot of the Graphical User Interface of SPPAS, showing how to perform the annotations.

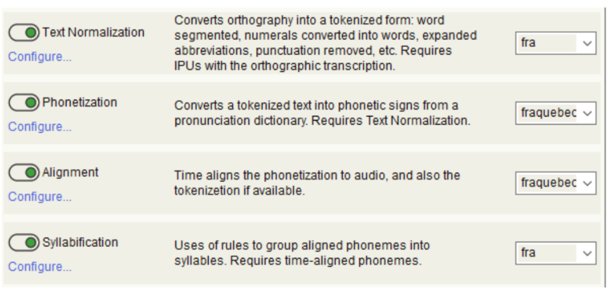

Figure 2: Screenshot of SPPAS automatic annotations

8. Conclusion

This paper presents the first automatic speech segmentation system freely available for Quebec French. It makes Text Normalization, Phonetization, Alignment and Syllabification automatic annotations available to perform a full speech segmentation process. Figure 3 illustrates the final result, compared to a manual one. The lexicon, the pronunciation dictionary and the first acoustic model are all freely distributed into SPPAS since version 2.7 (Nov. 2019). These newly created resources will be gradually improved and updated. This new release of SPPAS is, to our knowledge, the first freely available resource achieving a proper phonetization and segmentation for QF, taking into account the specific phonological structure of this variety as well as the local pronunciation of lexical items. It will allow phoneticians and phonologists to perform large scale data-based studies of QF with a reliable display of the proper phonemic symbols for each sound and an adapted speech segmentation. This is a significant contribution in and of itself, but also considering that French is spoken natively by more than seven million Canadians[3], making it the second largest community of speakers of French as a first language (after France).

[3] for details on our numbers see https://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-314-x/2011003/tbl/tbl3_1-1-fra.cfm

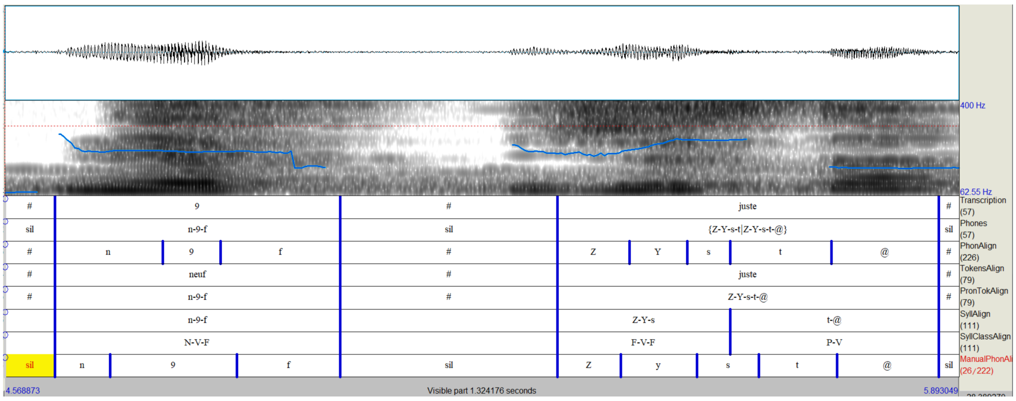

Figure 3: Illustration of the segmentation provided by SPPAS (3rd tier) versus manual segmentation (last tier).

9. Bibliographical References

Bigi, B. and Meunier, C. (2018). Automatic segmentation of spontaneous speech. *Revista de Estudos da Linguagem. International Thematic Issue: Speech Segmentation*, 26(4).

Bigi, B., Meunier, C., Nesterenko, I., and Bertrand, R. (2010). Automatic detection of syllable boundaries in spontaneous speech. In *7th International conference on Language Resources and Evaluation*, pages 3285–3292, La Valette, Malta.

Bigi, B., Péri, P., and Bertrand, R. (2012). Orthographic transcription: which enrichment is required for phonetization? In *Proceedings of the Eight International Conference on Language Resources and Evaluation*, pages 1756–1763, Istanbul, Turkey. European Language Resources Association (ELRA).

Bigi, B., Caron, B., and Oyelere, A. S. (2017). Developing resources for automated speech processing of the african language naija (nigerian pidgin). In *8th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 441–445, Poznań, Poland.

Bigi, B. (2014). A multilingual text normalization approach. *Human Language Technology Challenges for Computer Science and Linguistics*, LNAI-8387:515–526.

Bigi, B. (2015). SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, 111–112:54–69.

Bigi, B. (2016). A phonetization approach for the forced-alignment task in SPPAS. *Human Language Technology. Challenges for Computer Science and Linguistics*, LNAI-9561:397–410.

Boersma, P. and Weenink, D. (2018). Praat: doing phonetics by computer [computer program], version 6.0.37, retrieved 14 march 2018 from http://www.praat.org/.

Christodoulides, G. (2018). Praaline: An open-source system for managing, annotating, visualising and analysing speech corpora. In *Proceedings of ACL 2018, System Demonstrations*, pages 111–115.

Côté, M.-H. (2012). Laurentian french (quebec) extra vowels, missing schwas, and surprising liaison consonants. In *Phonological variation in French: Illustrations from three continents*, pages 235–274. Amsterdam, John Benjamins Publishing.

Côté, M.-H. (2014). Le projet pfc et la géophonologie du français laurentien. In *La phonologie du français. Normes, périphéries, modélisation*, pages 175–198. Presses Universitaires de Paris Ouest Nanterre.

Durand, J., Laks, B., and Lyche, C. (2002). La phonologie du français contemporain: usages, variétés et structure. *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language.*, pages 93–106.

Goldman, J.-P. (2011). Easyalign: An automatic phonetic alignment tool under praat. In *Interspeech*.

Klessa, K., Karpiński, M., and Wagner, A. (2013). Annotation pro-a new software tool for annotation of linguistic and paralinguistic features. In *Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop*, pages 51–54, Aix-en-Provence, France.

Le, V., Besacier, L., Seng, S., Bigi, B., and Do, T. (2008). Recent advances in automatic speech recognition for vietnamese. In *International Workshop on Spoken Languages Technologies for Under-resourced languages*, pages 47–52, Hanoi, Vietnam.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. In *Interspeech*, pages 498–502.

Santerre, L. (1976). Voyelles et consonnes du français québécois populaire. In *Identité culturelle et francophonie dans les Amériques*, volume 1, pages 21–36. PUL Québec.

Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. In *Proc. of the ICPhS*, pages 607–610, San Francisco, August.

Teston, B., Ghio, A., and Galindo, B. (1999). A multisensor data acquisition and processing system for speech production investigation. In *International Congress of Phonetic Sciences (ICPhS)*, pages 2251–2254. University of California.

Trimaille, C., Candea, M., and Lehka-Lemarchand, I. (2012). Existe-t-il une signification sociale stable et univoque de la palatalisation/affrication en français? étude sur la perception de variantes non standard. In *SHS Web of Conferences*, volume 1, pages 2249–2262. EDP Sciences.